



PCT10

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/049,750

DATE: 03/01/2002

TIME: 15:01:29

Input Set : A:\EP.txt

Output Set: N:\CRF3\03012002\J049750.raw

```

3 <110> APPLICANT: Tischer, Wilhelm
4   Ihlenfeldt, Hans-Georg
5   Barzu, Octavian
6   Sakamoto, Hiroshi
7   Pistotnik, Elisabeth
8   Marliere, Philippe
9   Pochet, Sylvie
12 <120> TITLE OF INVENTION: Enzymatic synthesis of deoxyribonucleosides
14 <130> FILE REFERENCE: 20373PWO Deoxyribonucleosides
C--> 16 <140> CURRENT APPLICATION NUMBER: US/10/049,750
17 <141> CURRENT FILING DATE: 2002-02-15
19 <150> PRIOR APPLICATION NUMBER: EP99116425.2
20 <151> PRIOR FILING DATE: 1999-08-20
22 <160> NUMBER OF SEQ ID NOS: 20
24 <170> SOFTWARE: PatentIn Ver. 2.1
26 <210> SEQ ID NO: 1
27 <211> LENGTH: 1323
28 <212> TYPE: DNA
29 <213> ORGANISM: Escherichia coli
31 <220> FEATURE:
32 <221> NAME/KEY: CDS
33 <222> LOCATION: (1)..(1320)
37 <400> SEQUENCE: 1
38 ttg ttt ctc gca caa gaa att att cgt aaa aaa cgt gat ggt cat gcg      48
39 Leu Phe Leu Ala Gln Glu Ile Ile Arg Lys Lys Arg Asp Gly His Ala
40   1           5           10          15
42 ctg agc gat gaa gaa att cgt ttc ttt atc aac ggt att cgc gac aac      96
43 Leu Ser Asp Glu Glu Ile Arg Phe Phe Ile Asn Gly Ile Arg Asp Asn
44   20           25           30
46 act atc tcc gaa ggg cag att gcc gcc ctc gcg atg acc att ttc ttc      144
47 Thr Ile Ser Glu Gly Gln Ile Ala Ala Leu Ala Met Thr Ile Phe Phe
48   35           40           45
51 cac gat atg aca atg cct gag cgt gtc tcg ctg acc atg gcg atg cga      192
52 His Asp Met Thr Met Pro Glu Arg Val Ser Leu Thr Met Ala Met Arg
53   50           55           60
55 gat tca gga acc gtt ctc gac tgg aaa agc ctg cat ctg aat ggc ccg      240
56 Asp Ser Gly Thr Val Leu Asp Trp Lys Ser Leu His Leu Asn Gly Pro
57  65           70           75           80
59 att gtt gat aaa cac tcc acc ggt ggc gtc ggc gat gtg act tcg ctg      288
60 Ile Val Asp Lys His Ser Thr Gly Gly Val Gly Asp Val Thr Ser Leu
61   85           90           95
63 atg ttg ggg ccg atg gtc gca gcc tgc ggc ggc tat att ccg atg atc      336
64 Met Leu Gly Pro Met Val Ala Ala Cys Gly Gly Tyr Ile Pro Met Ile

```

RAW SEQUENCE LISTING

DATE: 03/01/2002

PATENT APPLICATION: US/10/049,750

TIME: 15:01:29

Input Set : A:\EP.txt

Output Set: N:\CRF3\03012002\J049750.raw

```

65          100          105          110
67 tct ggt cgc ggc ctc ggt cat act ggc ggt acg ctc gac aaa ctg gaa 384
68 Ser Gly Arg Gly Leu Gly His Thr Gly Gly Thr Leu Asp Lys Leu Glu
69          115          120          125
72 tcc atc cct ggc ttc gac att ttc ccg gat gac aac cgt ttc cgc gaa 432
73 Ser Ile Pro Gly Phe Asp Ile Phe Pro Asp Asp Asn Arg Phe Arg Glu
74          130          135          140
76 att att aaa gac gtc ggc gtg gcg att atc ggt cag acc agt tca ctg 480
77 Ile Ile Lys Asp Val Gly Val Ala Ile Ile Gly Gln Thr Ser Ser Leu
78 145          150          155          160
80 gct ccg gct gat aaa cgt ttc tac gcg acc cgt gat att acc gca acc 528
81 Ala Pro Ala Asp Lys Arg Phe Tyr Ala Thr Arg Asp Ile Thr Ala Thr
82          165          170          175
84 gtg gac tcc atc ccg ctg atc acc gcc tct att ctg gcg aag aaa ctt 576
85 Val Asp Ser Ile Pro Leu Ile Thr Ala Ser Ile Leu Ala Lys Lys Leu
86          180          185          190
88 gcg gaa ggt ctg gac gcg ctg gtg atg gac gtg aaa gtg ggt agc ggc 624
89 Ala Glu Gly Leu Asp Ala Leu Val Met Asp Val Lys Val Gly Ser Gly
90          195          200          205
92 gcg ttt atg ccg acc tac gaa ctc tct gaa gcc ctt gcc gaa gcg att 672
93 Ala Phe Met Pro Thr Tyr Glu Leu Ser Glu Ala Leu Ala Glu Ala Ile
94          210          215          220
96 gtt ggc gtg gct aac ggc gct ggc gtg cgc acc acc gcg ctg ctc acc 720
97 Val Gly Val Ala Asn Gly Ala Gly Val Arg Thr Thr Ala Leu Leu Thr
98 225          230          235          240
100 gac atg aat cag gta ctg gcc tcc agt gca ggt aac gcg gtt gaa gtt 768
101 Asp Met Asn Gln Val Leu Ala Ser Ser Ala Gly Asn Ala Val Glu Val
102          245          250          255
106 cgt gaa gcg gtg cag ttc ctg acg ggt gaa tat cgt aac ccg cgt ctg 816
107 Arg Glu Ala Val Gln Phe Leu Thr Gly Glu Tyr Arg Asn Pro Arg Leu
108          260          265          270
110 ttt gat gtc acg atg gcg ctg tgc gtg gag atg ctg atc tcc ggc aaa 864
111 Phe Asp Val Thr Met Ala Leu Cys Val Glu Met Leu Ile Ser Gly Lys
112          275          280          285
114 ctg gcg aaa gat gac gcc gaa gcg cgc gcg aaa ttg cag gcg gtg ctg 912
115 Leu Ala Lys Asp Asp Ala Glu Ala Arg Ala Lys Leu Gln Ala Val Leu
116          290          295          300
118 gac aac ggt aaa gcg gca gaa gtc ttt ggt cgt atg gta gcg gca caa 960
119 Asp Asn Gly Lys Ala Ala Glu Val Phe Gly Arg Met Val Ala Ala Gln
120 305          310          315          320
122 aaa ggc ccg acc gac ttc gtt gag aac tac gcg aag tat ctg ccg aca 1008
123 Lys Gly Pro Thr Asp Phe Val Glu Asn Tyr Ala Lys Tyr Leu Pro Thr
124          325          330          335
126 gcg atg ctg acg aaa gca gtc tat gct gat acc gaa ggt ttt gtc agt 1056
127 Ala Met Leu Thr Lys Ala Val Tyr Ala Asp Thr Glu Gly Phe Val Ser
128          340          345          350
130 gaa atg gat acc cgc gcg ctg ggg atg gca gtg gtt gca atg ggc ggc 1104
131 Glu Met Asp Thr Arg Ala Leu Gly Met Ala Val Val Ala Met Gly Gly
132          355          360          365

```

RAW SEQUENCE LISTING

DATE: 03/01/2002

PATENT APPLICATION: US/10/049,750

TIME: 15:01:29

Input Set : A:\EP.txt

Output Set: N:\CRF3\03012002\J049750.raw

```

134 gga cgc cgt cag gca tct gac acc atc gat tac agc gtc ggc ttt act 1152
135 Gly Arg Arg Gln Ala Ser Asp Thr Ile Asp Tyr Ser Val Gly Phe Thr
136      370      375      380
140 gat atg gcg cgt ctg ggc gac cag gta gac ggt cag cgt ccg ctg gcg 1200
141 Asp Met Ala Arg Leu Gly Asp Gln Val Asp Gly Gln Arg Pro Leu Ala
142 385      390      395      400
144 gtt atc cac gcg aaa gac gaa aac aac tgg cag gaa gcg gcg aaa gcg 1248
145 Val Ile His Ala Lys Asp Glu Asn Asn Trp Gln Glu Ala Ala Lys Ala
146      405      410      415
148 gtg aaa' gcg gca att aaa ctt gcc gat aaa gca ccg gaa agc aca cca 1296
149 Val Lys Ala Ala Ile Lys Leu Ala Asp Lys Ala Pro Glu Ser Thr Pro
150      420      425      430
152 act gtc tat cgc cgt atc agc gaa taa 1323
153 Thr Val Tyr Arg Arg Ile Ser Glu
154      435      440
157 <210> SEQ ID NO: 2
158 <211> LENGTH: 440
159 <212> TYPE: PRT
160 <213> ORGANISM: Escherichia coli
162 <400> SEQUENCE: 2
163 Leu Phe Leu Ala Gln Glu Ile Ile Arg Lys Lys Arg Asp Gly His Ala
164 1      5      10      15
166 Leu Ser Asp Glu Glu Ile Arg Phe Phe Ile Asn Gly Ile Arg Asp Asn
167      20      25      30
169 Thr Ile Ser Glu Gly Gln Ile Ala Ala Leu Ala Met Thr Ile Phe Phe
170      35      40      45
174 His Asp Met Thr Met Pro Glu Arg Val Ser Leu Thr Met Ala Met Arg
175      50      55      60
177 Asp Ser Gly Thr Val Leu Asp Trp Lys Ser Leu His Leu Asn Gly Pro
178 65      70      75      80
180 Ile Val Asp Lys His Ser Thr Gly Gly Val Gly Asp Val Thr Ser Leu
181      85      90      95
183 Met Leu Gly Pro Met Val Ala Ala Cys Gly Gly Tyr Ile Pro Met Ile
184      100      105      110
186 Ser Gly Arg Gly Leu Gly His Thr Gly Gly Thr Leu Asp Lys Leu Glu
187      115      120      125
189 Ser Ile Pro Gly Phe Asp Ile Phe Pro Asp Asp Asn Arg Phe Arg Glu
190      130      135      140
192 Ile Ile Lys Asp Val Gly Val Ala Ile Ile Gly Gln Thr Ser Ser Leu
193 145      150      155      160
195 Ala Pro Ala Asp Lys Arg Phe Tyr Ala Thr Arg Asp Ile Thr Ala Thr
196      165      170      175
198 Val Asp Ser Ile Pro Leu Ile Thr Ala Ser Ile Leu Ala Lys Lys Leu
199      180      185      190
201 Ala Glu Gly Leu Asp Ala Leu Val Met Asp Val Lys Val Gly Ser Gly
202      195      200      205
204 Ala Phe Met Pro Thr Tyr Glu Leu Ser Glu Ala Leu Ala Glu Ala Ile
205      210      215      220
209 Val Gly Val Ala Asn Gly Ala Gly Val Arg Thr Thr Ala Leu Leu Thr

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/049,750

DATE: 03/01/2002

TIME: 15:01:29

Input Set : A:\EP.txt

Output Set: N:\CRF3\03012002\J049750.raw

```

210 225          230          235          240
212 Asp Met Asn Gln Val Leu Ala Ser Ser Ala Gly Asn Ala Val Glu Val
213          245          250          255
215 Arg Glu Ala Val Gln Phe Leu Thr Gly Glu Tyr Arg Asn Pro Arg Leu
216          260          265          270
218 Phe Asp Val Thr Met Ala Leu Cys Val Glu Met Leu Ile Ser Gly Lys
219          275          280          285
221 Leu Ala Lys Asp Asp Ala Glu Ala Arg Ala Lys Leu Gln Ala Val Leu
222          290          295          300
224 Asp Asn Gly Lys Ala Ala Glu Val Phe Gly Arg Met Val Ala Ala Gln
225 305          310          315          320
227 Lys Gly Pro Thr Asp Phe Val Glu Asn Tyr Ala Lys Tyr Leu Pro Thr
228          325          330          335
230 Ala Met Leu Thr Lys Ala Val Tyr Ala Asp Thr Glu Gly Phe Val Ser
231          340          345          350
233 Glu Met Asp Thr Arg Ala Leu Gly Met Ala Val Val Ala Met Gly Gly
234          355          360          365
236 Gly Arg Arg Gln Ala Ser Asp Thr Ile Asp Tyr Ser Val Gly Phe Thr
237          370          375          380
239 Asp Met Ala Arg Leu Gly Asp Gln Val Asp Gly Gln Arg Pro Leu Ala
240 385          390          395          400
244 Val Ile His Ala Lys Asp Glu Asn Asn Trp Gln Glu Ala Ala Lys Ala
245          405          410          415
247 Val Lys Ala Ala Ile Lys Leu Ala Asp Lys Ala Pro Glu Ser Thr Pro
248          420          425          430
250 Thr Val Tyr Arg Arg Ile Ser Glu
251          435          440
255 <210> SEQ ID NO: 3
256 <211> LENGTH: 720
257 <212> TYPE: DNA
258 <213> ORGANISM: Escherichia coli
260 <220> FEATURE:
261 <221> NAME/KEY: CDS
262 <222> LOCATION: (1)..(717)
264 <400> SEQUENCE: 3
265 atg gct acc cca cac att aat gca gaa atg ggc gat ttc gct gac gta 48
266 Met Ala Thr Pro His Ile Asn Ala Glu Met Gly Asp Phe Ala Asp Val
267 1 5 10 15
269 gtt ttg atg cca ggc gac ccg ctg cgt gcg aag tat att gct gaa act 96
270 Val Leu Met Pro Gly Asp Pro Leu Arg Ala Lys Tyr Ile Ala Glu Thr
271 20 25 30
273 ttc ctt gaa gat gcc cgt gaa gtg aac aac gtt cgc ggt atg ctg ggc 144
274 Phe Leu Glu Asp Ala Arg Glu Val Asn Asn Val Arg Gly Met Leu Gly
275 35 40 45
279 ttc acc ggt act tac aaa ggc cgc aaa att tcc gta atg ggt cac ggt 192
280 Phe Thr Gly Thr Tyr Lys Gly Arg Lys Ile Ser Val Met Gly His Gly
281 50 55 60
283 atg ggt atc ccg tcc tgc tcc atc tac acc aaa gaa ctg atc acc gat 240
284 Met Gly Ile Pro Ser Cys Ser Ile Tyr Thr Lys Glu Leu Ile Thr Asp

```

RAW SEQUENCE LISTING

DATE: 03/01/2002

PATENT APPLICATION: US/10/049,750

TIME: 15:01:29

Input Set : A:\EP.txt

Output Set: N:\CRF3\03012002\J049750.raw

```

285 65          70          75          80
287 ttc ggc gtg aag aaa att atc cgc gtg ggt tcc tgt ggc gca gtt ctg 288
288 Phe Gly Val Lys Lys Ile Ile Arg Val Gly Ser Cys Gly Ala Val Leu
289          85          90          95
291 ccg cac gta aaa ctg cgc gac gtc gtt atc ggt atg ggt gcc tgc acc 336
292 Pro His Val Lys Leu Arg Asp Val Val Ile Gly Met Gly Ala Cys Thr
293          100          105          110
295 gat tcc aaa gtt aac cgc atc cgt ttt aaa gac cat gac ttt gcc gct 384
296 Asp Ser Lys Val Asn Arg Ile Arg Phe Lys Asp His Asp Phe Ala Ala
297          115          120          125
299 atc gct gac ttc gac atg gtg cgt aac gca gta gat gca gct aaa gca 432
300 Ile Ala Asp Phe Asp Met Val Arg Asn Ala Val Asp Ala Ala Lys Ala
301          130          135          140
303 ctg ggt att gat gct cgc gtg ggt aac ctg ttc tcc gct gac ctg ttc 480
304 Leu Gly Ile Asp Ala Arg Val Gly Asn Leu Phe Ser Ala Asp Leu Phe
305          145          150          155          160
307 tac tct ccg gac ggc gaa atg ttc gac gtg atg gaa aaa tac ggc att 528
308 Tyr Ser Pro Asp Gly Glu Met Phe Asp Val Met Glu Lys Tyr Gly Ile
309          165          170          175
313 ctc ggc gtg gaa atg gaa gcg gct ggt atc tac ggc gtc gct gca gaa 576
314 Leu Gly Val Glu Met Glu Ala Ala Gly Ile Tyr Gly Val Ala Ala Glu
315          180          185          190
317 ttt ggc gcg aaa gcc ctg acc atc tgc acc gta tct gac cac atc cgc 624
318 Phe Gly Ala Lys Ala Leu Thr Ile Cys Thr Val Ser Asp His Ile Arg
319          195          200          205
321 act cac gag cag acc act gcc gct gag cgt cag act acc ttc aac gac 672
322 Thr His Glu Gln Thr Thr Ala Ala Glu Arg Gln Thr Thr Phe Asn Asp
323          210          215          220
325 atg atc aaa atc gca ctg gaa tcc gtt ctg ctg ggc gat aaa gag taa 720
326 Met Ile Lys Ile Ala Leu Glu Ser Val Leu Leu Gly Asp Lys Glu
327          225          230          235
330 <210> SEQ ID NO: 4
331 <211> LENGTH: 239
332 <212> TYPE: PRT
333 <213> ORGANISM: Escherichia coli
335 <400> SEQUENCE: 4
336 Met Ala Thr Pro His Ile Asn Ala Glu Met Gly Asp Phe Ala Asp Val
337 1 5 10 15
339 Val Leu Met Pro Gly Asp Pro Leu Arg Ala Lys Tyr Ile Ala Glu Thr
340 20 25 30
342 Phe Leu Glu Asp Ala Arg Glu Val Asn Asn Val Arg Gly Met Leu Gly
343 35 40 45
347 Phe Thr Gly Thr Tyr Lys Gly Arg Lys Ile Ser Val Met Gly His Gly
348 50 55 60
350 Met Gly Ile Pro Ser Cys Ser Ile Tyr Thr Lys Glu Leu Ile Thr Asp
351 65 70 75 80
353 Phe Gly Val Lys Lys Ile Ile Arg Val Gly Ser Cys Gly Ala Val Leu
354 85 90 95
356 Pro His Val Lys Leu Arg Asp Val Val Ile Gly Met Gly Ala Cys Thr

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/049,750

DATE: 03/01/2002

TIME: 15:01:30

Input Set : A:\EP.txt

Output Set: N:\CRF3\03012002\J049750.raw

L:16 M:270 C: Current Application Number differs, Replaced Current Application Number